DOCUMENT RESUME

ED 177 209 TM 009 716

AUTHOR Olejnik, Stephen F.
TITLE Standardized Achievement Testing Programs Viewed From the Perspective of a Non-Measurement Specialist.
PUB DATE Apr 79
NOTE 24p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April 9-11, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; Administrator Attitudes; Counselor Attitudes; Elementary Secondary Education; Information Utilization; Knowledge Level; Standardized Tests; *Student Testing; *Teacher Attitudes; *Testing Problems; *Test Interpretation; *Test Results

ABSTRACT

Despite a decade of debate, standardized achievement testing continues as one of the most controversial issues in education. This study sought the views of teachers, counselors, and principals, regarding the advantages and limitations of achievement testing programs. Responses to a questionnaire and informal discussions indicated positive attitudes toward testing, but little usage beyond the elementary school level. Results also indicated the educators were ill-prepared to interpret test scores, but were eager to learn. Finally, the need to increase student motivation and the difficulty of the reading level of examinations were cited as major concerns. (Author/MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED177209

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Standardized Achievement Testing Programs Viewed From The Perspective of the Non-Measurement Specialist

Stephen F. Olejnik

University of Florida

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Stephen F. Olejnik

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of
the National Council on Measurement in Education
San Francisco, California April, 1979

TM 009 716

ABSTRACT

In spite of over a decade of debate, standardized acheivement testing continues as one of the most controversial issues in education. Non-measurement specialists have been excluded from these discussions. The reported study sought the views of teachers, counselors, and principals, regarding the advantages and limitations of achievement testing programs. Responses to a questionnaire and informal discussions indicated positive attitudes toward testing, but little usage beyond the elementary school level. Results also indicated the educators were ill-prepared to interpret test scores, but were eager to learn. Finally, increasing student motivation and the reading level of exams were cited as major concerns.

Standardized Achievement Testing Programs

Viewed from the Perspective of the Non-Measurement Specialist

Since the turn of the century, psychological testing has developed into a multi-million dollar industry. Its growth can be attributed, at least in part, to the belief that standardized test results provide useful information for making decisions regarding individuals. Educators in particular have become major consumers of the products produced by the testing industry. Standardized tests including intelligence, aptitude, achievement, personality and interest inventories are administered to millions of students each year. The National Educational Association estimates that at least 200 million achievement test forms alone are used each year in the United States (Principal, 1975). The increased "popularity" of psychological testing in schools has also brought considerable controversy concerning the desirability of these tests. While all types of tests have been criticized, instruments designed to measure intelligence and achievement have received the greatest attention. The present investigation concentrated on the issue of standardized achievement testing.

Standardized achievement tests were originally introduced to provide teachers with an objective method of monitoring student educational progress. An individual's progress was to be judged in relation to the progress made by the student's peers, both nationally and locally. Testing advocates were quick to suggest, however, that these instruments could be used for several additional purposes including: the evaluation of new and traditional programs, as part of an overall teacher accountability system and as a basis for certifying student competencies. Since achievement tests seemed to offer something for just about

everyone, they were adopted by many school districts. Thus, testing programs developed and expanded from a single achievement test to a test battery, from testing at selected grade levels to testing in all grades, and from a single testing session to Fall and Spring testing and more.

While standardized achievement tests grew in "popularity", its growth was not without controversy. Almost from the very beginning of the testing movement, critics have voiced their objections to the use of achievement tests in schools. These instruments were criticized for errors in logic, item ambiguity and for being biased against some groups in society. Test opponents also challenged the multiple choice format as being inadequate in measuring important educational outcomes such as the thought process which goes into selecting an answer. Still another argument raised by test critics has been with regard to the alleged influence tests have on the school curriculum. Some have charged that obtaining good test results has become an end in itself. As a result of the concern for high test performance, teachers may be hesitant to try new innovative instructional methods. Thus, the tests are seen as a major force of dictating classroom curricula (These and other objections were raised by tests critics in the July-August, 1975, issue of the National Elementary Principal). Arguments such as these have resulted in the reduction or re-evaluation of many testing programs.

Although a great deal has been said regarding the advantages and limitations associated with achievement testing programs, most of the comments have been made by individuals who were not directly involved with schools or classrooms on a daily basis. Very few studies (Goslin, 1967; Jackson, 1968) have examined the issue of how standardized achievement test results are used within a school system. Do the advantages

of a testing program really exist when they are translated from the ideal as conceptualized by measurement experts in the testing industry to the real life setting of the local school district? Or do well-meaning school personnel ignore, misuse or abuse the test results as has been charged by test critics. The present study attempted to address these issues by questioning a group of non-measurement experts with regard to their attitudes toward achievement testing, their use of the test results and their preparation/ability to interpret the test results. The definition of a non-measurement expert suggested by Stetz and Beck (1978) as "those directly involved with educational tests and measurements who do not have a degree related to this field of specialization", was adopted. For this study, non-measurement experts included school administrators, counselors, and teachers from a single school district.

Participating District

A single criterion was set in selecting the school district for the study and it was simply that an extensive achievement testing program had to be in operation. The district that was chosen administered the Stanford Achievement Test to all students in grades 2,3,5,7 and 8 and the Stanford Test of Academic Skills to all students in grades 9, 10, 11 and 12. All of these tests were given in the Fall of each academic year.

The student popualtion consisted of approximately 8500 pupils enrolled in six elementary schools, two middle schools and a single high school. Families living within the district differed considerably in terms of their social-economic backgrounds and the schools within the system reflected this diversity. Two of the elementary schools were located in newly-developed portions of the district and had students primarily from middle or upper-middle class backgrounds. Two other elementary schools were located in older sections of the district

and their students came primarily from working class or blue-collar homes. Two of the six elementary schools were located in sections of the district where many apartments had been recently built. These schools provided instruction to children whose families were considerably more mobile than other families in the district. It was believed that if responses to questions on testing attitude and test use varied, the variation might be explained in terms of the diverse student population enrolled in the schools. The middle schools and the high school provided instruction to students coming from a cross section of SES backgrounds.

Sample

A total of 200 questionnaires were distributed to all of the administrators (principals and assistant principals), all of the counselors and a sample of teachers from all public schools in the district. At the elementary school level, teachers of grades 3, 4, and 5 were asked to respond to the instrument. At the middle school level, teachers of grades 6, 7, and 8 were given the questionnaire. At the high school level, only teachers of language arts or mathematics in grades 9, 10, 11, and 12 were asked to complete the questionnaire. Sixty-nine percent or 137 of the questionnaires were returned. Only one elementary school refused to participate in the survey.

The responding sample consisted of 86% teachers, 7% counselors and 7% administrators. Forty percent were affiliated with the elementary schools, 37% were affiliated with the middle schools and 23% were affiliated with the high school. In terms of educational backgrounds: 14% of the respondents were certified with a BA/BS degree, 36% had taken graduate level courses beyond the bachelor's degree and 50% of the respondents had a Master's degree.

Results

The items on the questionnaire were classified into three broad categories: 1) attitudes toward standardized achievement testing, 2) use of the test results and 3) the preparation/ability of respondents to properly interpret the test results. Responses to each item within these categories were analyzed in terms of the educational backgrounds of the respondents, characteristics of the community, position held by the respondents and the school level at which the respondents worked. Only the latter analysis appeared to provide significant variation in the pattern of responses. These results are reported below. For each item, the percent of the total sample and the percent of the respondents at each school level selecting each foil is presented.

Attitudes

Five items on the questionnaire were designed to obtain attitudinal information concerning the achievement testing program in the district. Tables 1 through 5 present these items and the percent of the respondents selecting each foil. Tables 1 and 2 report the most direct measure of attitude toward the testing program. Over two-thirds of the respondents felt that the information provided by the tests was worth the time spent away from the classroom. Only 11% of the total sample indicated that they felt it would be better not to test the students at all. Tables 3 and 4 present the responses to items concerning the perceived importance of the test results. These items may be conceived of as indirect indicies of attitudes toward the testing program. An individual having a negative attitude toward the tests would likely view test results as having little or no importance. Most of the respondents, however, seemed to feel that it was at least somewhat important that students do well on the test. It was somewhat important that as educators

Table 1

Do you think the information provided by test results is worth the time spent away from the classroom?

	Total Sample	Elementary	Middle	High
Yes.	68	75	63	66
No.	32	25	37	34

Table 2

When do you think it is better to test students?

	Total Sample	Elementary	Middle	High
At the beginning of the school year.	20	32	8	17
At the end of the school year	22	26	24	10
Both at the beginning and end of the school year	47	35	56	55
Neither at the beginning or the end of the school year	11	7	12	17

Table 3

Is it important to you that the students do well on the Stanford Achievement Test?

	Total Sample	Elementary	Middle	High
Very Important	12	6	16	16
Some Importance	62	70	61	50
Little Importance	20	20	14	28
No Importance	7	4	10	6

Table 4

Is it important for you to know how well the students scored on the Stanford Achievement Test?

	Total Sample	Elementary	Middle	High
Very Important	14	17	14	9
Somewhat Important	61	63	57	67
Not Important	25	20	29	25

Table 5

Below are several scources of information about student abilities. Please rank these sources by assigning a rank of one (1) to the most important (reliable and valid), a rank of two (2) to the next most important and so forth.

	Total Sample	Elementary	Middle	High
Previous grades	42	44	45	41
Teacher comments	29	34	33	23
Counselor comments	4	2	6	3
IQ scores	10	9	8	16
Standardized Ach. test results	5	9	4	0
Family background	6	4	4	16

they be aware of how well the students performed. Table 5 presents the responses to an item in which the educators were asked to rank several sources of information on student abilities. The entries of the table are the percent of the respondents ranking each option as the most important (reliable and valid) source of information on student abilities. Only a small percentage of the respondents chose either achievement test scores or IQ scores as being the most important.

In general, it appears that the non-measurement specialists in the district surveyed have a positive attitude toward the achievement testing program. These results seem consistent across the school levels: elementary, middle and high. When achievement test scores were compared to other sources of information on student abilities, most of the non-measurement specialists felt that the test scores were of less importance (less reliable and valid). These results seem to be inconsistent with the positive attitudes expressed in the other items. In comments made during the informal interviews, many educators expressed concerns which may explain the low rankings of the achievement test results. Many respondents expressed skepticism regarding student performance because of several factors: 1) students didn't see the tests as being important, 2) the reading level of the tests was too high for many students and they lost interest; and 3) the tests took too long to take and students lost their motivation to perform well. Thus, while educator attitudes were positive, student attitudes may be negative, resulting in low student performances which are unreliable and invalid.

Using Test Results

The positive attitudes expressed by the non-measurement specialists in the previous section provides one index of how local educators view standardized achievement tests. An even more important indication, however, of the value placed on these measures may be found in the

extent to which the test results are used in the daily operation of the schools and classrooms. Nine items on the questionnaire addressed this issue. These items and the percent of the respondents selecting each of the foils are presented in Tables 6 through 14. In contrast to the items measuring attitude, the responses to questions on usage varied considerably across the school levels.

Table 6 presents the percent of the respondents who indicated easy access to the test results. While this item does not concern use of the test results, the response pattern to the question does provide information useful in explaining other items considered in this section. Eighty-seven percent of the total sample indicated that they had easy access to the test results of their students. This result is misleading, however, since the percent of educators with easy access to the scores declines steadily as the school level increases. At the high school level, 30% of the respondents indicated that they did not have easy access to the test results. In order to make a reasonable comparison of the response patterns across the school levels in terms of test usage (tables 7 through 13) only those respondents indicating easy access to the test files were considered.

Table 7 presents the response pattern to an item which asked the educators the extent to which they were familiar with the specific objectives of the achievement measure taken by their students. At the elementary and middle school levels almost two-thirds of the respondents indicated that they were familiar with the objectives. But at the high school level only a little more than one-third of the respondents indicated that they had knowledge of the test objectives. It seems unlikely that the educators at the high school level could use the test results in any meaningful way if they were unaware of what the test was measuring.

Table 6

Do you have easy access to the files containing student test results?

	Total Sample	Elementary	Middle	High
Yes, of all students in the school	67	70	64	66
Yes, but only my present students	20	30	20	3
No.	13	0	16	30

Table 7

Are you familiar with the specific objectives tested by the Stanford Achievement Test for your grade level?

	Elementary	Middle	High
Yes.	65	60	38
No.	35	40	62

Table 8

Some school personnel say they never look at test results. Do you examine the test results of individual students?

	Elementary	Middle	High
Yes, always	57	16	14
Yes, sometimes	41	67	62
No.	2	16	23

Table 9

Do you examine the test results of your classroom(s) as a group?

	Elementary	Middle	High
Yes, always	52	16	5
Yes, sometimes	35	44	29
No.	13	40	67

The non-measurement specialists were asked to what extent the test files were examined and when during the year were the scores looked at. Tables 8,9 and 10 present the response patterns to these items. The results indicate that only educators at the elementary level examine the test files to any great extent. Neither the middle or high school educators seem to bother with the test scores although both groups had indicated easy access to the test files. Table 10 presents the responses to an item which asked when during the year were the results examined (more than one foil could be selected). Very few of the educators in any of the school levels indicated that they examined the results from previous year's testing. It seems unlikely then that tests play an important role in planning instruction during the early part of the school year. Students in the district were given the test in early fall and the scores were reported to the teachers in early December. Almost two-thirds of the educators at the elementary level indicated that they examined these results when they were first reported. At the middle and high school levels however less than one-third of the respondents indicated that they looked at the test results when they were first reported. Finally almost half of the elementary and high school educators who had easy access to the test results indicated that they would look at the results periodically throughout the school year. Only one-third of the middle school educators indicated this practice.

The educators were asked the extent to which the test results influenced their instructional planning; the extent to which the results influenced what was taught in the classroom; and the extent to which they felt the test results influenced educational decision making. The results of these items are presented in tables 11, 12 and 13 respectively. At the elementary and middle school levels the test results seem to have some influence in instructional planning. These results seem inconsistent with the response patterns presented

Table 10

If you do examine the test results when during the year do you do so?

	Elementary	Middle	High
Test results from previous years are examined within the first couple of weeks of the school year.	35	19	5
When the current year test results are first reported	65	26	24
Periodically throughout the year.	46	33	48

Table 11

Do test results influence your instructional planning for individual students?

	Elementary	Middle	High
Yes, always.	6	12	5
Yes, occasionally	48	40	29
Yes, but rarely	25	21	19
No.	21	28	48

Table 12

What influence does the testing program have on what is taught in the classroom?

	Elementary	Middle	High
A great deal of influence	2	2	3
Some influence	40	31	33
Little influence	47	52	38
No influence	11	14	29

in tables 8, 9 and 10 for the middle school educators. At the high school the educators indicated that the test results were of little influence in instructional planning.

Educators across the three school levels indicated that the test results had little influence on what they taught in the classroom. It therefore seems that for this district the test did not dictate the classroom curriculum.

Almost two-thirds of the elementary and middle school educators indicated that they felt that the test results had some influence on educational decision making. At the high school level the educators indicated the opposite view.

Finally concerning the use of test results, the educators were asked what form of reported test scores were most useful to them. Responses from the total sample to this item are presented in table 14. The educators across the three school levels seemed to agree that percentiles and grade equivalents were the most useful forms of reported scores. The responses to this item were somewhat surprising in two ways. First a considerable percentage of the educators indicated that the raw scores were the most useful. And second very few respondents chose stanines as being the most useful. This is surprising since the test publisher recommends stanines and uses stanine scores in its examples of how the test results could/should be interpreted. The choice of grade equivalents as being the most useful form of reported scores disheartening when one considers the repeated efforts of test experts to discourage their use.

Comments made during the informal interviews regarding test use included such issues as: a) slow returns of the testing results, b) incomplete feedback to teachers concerning the test results and c) a lack of teacher time to examine and interpret the test results. Finally many educators indicated that they used the test results only when they had problems with student learning and wanted to know the reason.

Table 13

What influence do you think the testing program plays in educational decision making?

	Elementary	Middle	High
A major influence	8	5	10
Some influence	64	67	38
Little influence	24	28	38
No influence	4	0	15

Table 14

Which form of reported scores are most useful to you?

	Total Sample	Elementary	Middle	High
Raw scores	13	11	20	6
Stanines	12	19	6	13
Percentiles	45	46	39	53
Scaled Scores	2	2	0	6
Grade Equivalents	52	69	51	25

Interpreting Test Results

The third category which the questionnaire items were classified was the area of educator preparation/ability in properly interpreting test results. When the study began it was felt that if the non-measurement specialists were not using the test results it might be because the educators lacked the training or confidence in their ability to interpret the test scores. Three items on the questionnaire reflected respondent training and confidence in test interpretation. These items and the response patterns are presented in tables 15, 16 and 17.

Almost two-thirds (table 15) of the responsents indicated that they had taken a course as an undergraduate or graduate which dealt with interpreting test results. But only at the elementary level did the educators indicate that they had used or seen the publishers test manual. Finally almost 20% of the middle and high school level educators indicated that they had no training in test interpretation.

Over 90% of the elementary and middle school educators indicated that they were at least somewhat confident in their ability to interpret test scores (table 16). At the high school level however over 40% of the educators indicated that they were not at all confident in their test interpretation skills. In the previous section the results had indicated that among the three school levels the respondents from the high school used the test results the least. While the confidence to correctly interpret test results varied across the school levels, almost three-fourths of the educators in the sample agreed that additional assistance was needed in test interpretation (table 17).

The previous three questions were followed by five items which asked the respondents to interpret some test scores. These items were similar to those which might be asked in an elementary course on tests and measurement. Responses

Table 15

What training have you had in interpreting test results?

	Total Sample	Elementary	Middle	High
No training	14	7	18	19
The test publisher's manual	47	74	31	28
Some inservice training	18	26	18	3
Some undergraduate/graduate courses	64	65	65	59

Table 16

How confident do you feel in correctly interpreting test results?

	Total Sample	Elementary	Middle	High
Very confident	17	22	14	13
Somewhat confident	69	72	78	47
Not confident at all	15	6	8	41

Table 17

Do you think additional assistance should be provided in interpreting the test results?

	Total Sample	Elementary	Middle	High
Yes.	72	64	78	74
No.	28	36	22	26

to these items are presented in tables 18, 19, 20, 21 and 22. The non-measurement specialists were consistent across school levels in their choices of item foils. Most of the respondents correctly answered the item regarding a percentile score (table 18). Over three-fourths of the educators indicated that they are incorrectly interpreting norms as standards of performance (table 19). Less than half of the elementary and about one fourth to the middle and high school educators indicated knowledge of how to interpret stanines (table 21). And finally while the educators had indicated that a grade equivalent score was a very useful form of reported results, only a small percent of the educators correctly answered the item on grade equivalent scores correctly.

Conclusions

The results of the survey seem to indicate that in general non-measurement specialists across the elementary, middle and high school levels view standardized achievement tests positively. The responses however were not verified with any other data source. Educators may have responded not in terms of their actual feelings but rather in ways which may be viewed as being porfessionally desirable. Assuming that the responses do reflect the attitudes of non-measurement specialists, these results appear to support the perceptions of superintendents reported by Stetz and Beck (1978) rather than the perceptions of NCME members.

It was also suggested that the extent to which educators use test results may be a better index of how non-measurement specialists view standardized achievement tests. If this is true then only elementary educators view these instruments positively. Middle and high school educators indicated that they made very little use of the test results. These results were obtained after taking into consideration the accessibility of the test files.

Table 18

If John received a score on the math concepts subtest equalling the 75th percentile, it would mean:

	Total Sample	Elementary	Middle	High
He answered 75% of the items correctly	2	0	2	3
He answered 75% of the items incorrectly	0	0	0	0
75% of all pupils in the norm group scored as well or better than John	15	20	13	10
*75% of all pupils in the norm group scored as well or lower than John	73	69	77	74
None of the above is correct.	10	11	6	13

Table 19

Normative data like grade equivalent scores provide a standard which reflects an estimate of how well students should perform.

	Total Sample	Elementary	Middle	High
True	77	80	80	67
*False	23	20	20	33

Table 20

To compare student performance across subtests which form of reported scores is most useful?

	Total Sample	Elementary	Middle	High
Raw scores	14	8	15	24
*Percentiles	24	26	26	20
Grade Equivalents	25	27	25	20
Scaled Scores	10	8	17	0
All of the above	24	29	17	28
None of the above	2	2	0	8

Table 21

The difference in student performance on two subtests on the Stanford Achievement Test should be judged meaningful if one test is at least _____ lower than the other.

	Total Sample	Elementary	Middle	High
1 stanine	15	7	22	20
*2 stanines	35	46	27	25
3 stanines	14	16	14	10
4 stanines	4	7	0	5
Stanines do not provide a useful scale to compare subtests	33	25	38	40

Table 22

Sally is finishing the 3rd grade. On a recent administration of the Stanford Achievement Test her total test battery score indicated a grade equivalent of 5.0. An appropriate interpretation would be:

	Total Sample	Elementary	Middle	High
Sally is ready to begin the fifth grade.	7	6	7	11
Sally knows as much as the average student beginning the fifth grade	61	57	63	67
*Compared to beginning fifth graders in the norm group Sally is at the 50th percentile	12	11	17	4
None of the above is correct	20	26	13	19

A partial explanation for the minimal use of the test results may be found in the training/ability of the educators to properly interpret the test results. While most of the respondents had taken course work related to test interpretation and felt somewhat confident in these skills, almost three-fourths of the educators indicated that they would like additional assistance in test interpretation. Based on the responses to several items asking the educators to interpret several test scores, it appears that the non-measurement specialists need the additional assistance.

References

Goslin, D.A. Teachers and Testing, New York: Russell Sage Foundation 1967.

Jackson, P. Life in the Classroom, New York: Holt, Rinehart and Winston Inc. 1968.

Stetz, and Beck, M.,"A survey of opinions concerning users of educational tests", Paper presented at the meeting of the National Council on Measurement in Education, Toronto, Ontario, 1978.

The Scoring of Children: Standardized Testing in America , National Elementary Principal 54 July/August 1975.